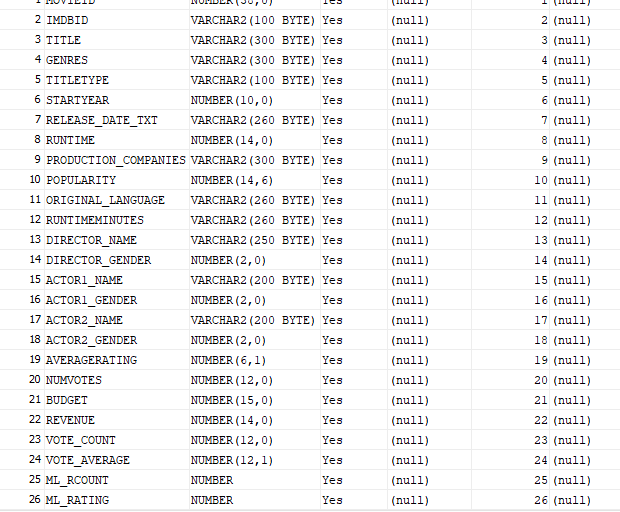
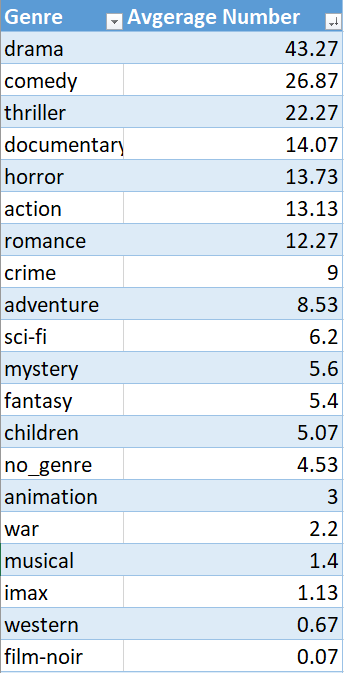
1. Joined and created the new table – movie with row count: 17071
2. Total 26 variables



* 1. ---1. What are the average number of English speaking films released each year over the most recent 15 years?

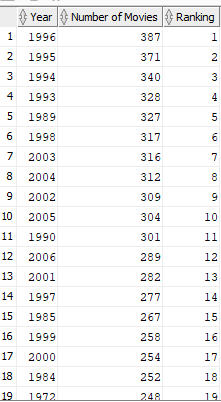


---2. Categorize #1 by genre.



Add up = 198.41

---3. What year had the largest number of films? By Genre? (Rank them?)



---4. **Can you report by gender of the Director? (Hint: OAC can enrich data by determining gender from name or maybe there is a column that will help)**

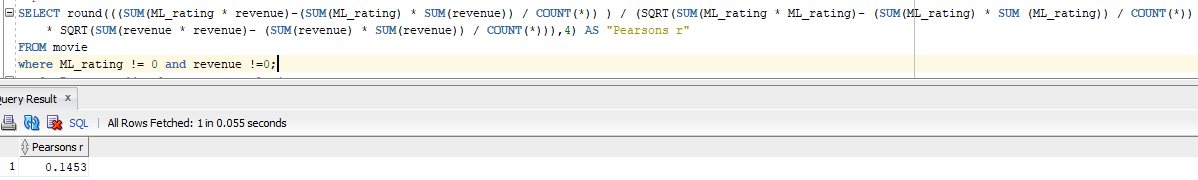
**---5. What are the top Directors? Actor? Male or Female top Actors?**

**---6. What are the top Rated and bottom rated films per year and overall?**

---7. Do the same by revenue.

---8. Do Ratings directly correlate with Revenue?

Pearson’s Product-moment Correlation Coefficient gives a measurement from -1 for a perfect negative correlation (as one variable goes up, the other goes down) to 1 for a perfect correlation (as one variable goes up, the other goes up). a correlation of 0 means that there is no relationship between the two. Pearson’s correlation coefficient is calculated as the covariance of the two variables divided by the product of their standard deviations.





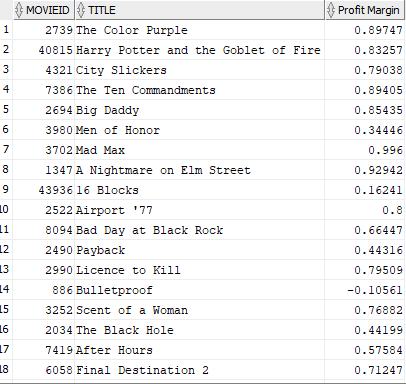
---9**. Do tags add value to any analysis?**

**---10. Do Actors add value to the analysis?**

**---11. Do Directors add value to the analysis?**

**---12. Are there any insights that can be made by using Production Company?**

---13. Can you determine profitability of the movies?



2338 of them have both revenue and budget

1. Transform the data to new dataset called movie\_new
   1. Exclude useless data
   2. Log transformed extreme value
   3. Convert the genres to DTM :
      1. A screen shot of a computer

         Description automatically generated
      2. Total 19 genres captured
      3. A screen shot of a computer

         Description automatically generated file\_name: genres\_summary
2. Created TAG table, and scripts called “tag\_creation”. It includes movies id, tag, and relevance score for further analysis
3. 1